基于有向合著网络的核心团队挖掘研究*
——以图情领域为例

高楠[1,2] 周庆山[2]
[1] 中国科学技术信息研究所 北京 100038 [2]北京大学信息管理系 北京 100871

摘 要：[目的/意义]为便于机构人才引进，构建领域核心团队识别规范化流程，从核心区域作者确定、团队指导/师承关系判别、团队发刊倾向和团队研究方向 4 个方面对领域核心团队进行挖掘研究。[方法/过程]首先，进行数据前期规范；然后，结合作者合著无向二值矩阵与核心－边缘结构分析，划分核心区域作者，综合多维指标对核心区域作者进行适当调整；最后，基于第一作者与其他作者的有向合著网络，识别领域核心研究团队。[结果/结论]在厘清团队指导/师承关系判别基本流程上，提炼通用性规律，包括：①在“外射星型”拓扑结构的合著团队中，处于中心位置节点有较高概率为导师/指导者，而“内聚星型”拓扑结构合著团队位于网络中心的作者科研能力较强；②在第一作者与其他作者的有向合著网络中，两个节点若同时具备第一作者与通讯作者关系、有相同的所属机构、学术年龄有一定差距，则这两个节点间一定存在指导关系，有较高的概率存在师承关系。

关键词：社会网络分析 有向合著网络 合著 核心－边缘结构
分类号：G250
DOI：10.13266/j.issn.0252－3116.2021.20.009

1 引言

核心团队是开展科学技术研究的基本组成单元，是由优势互补、联系紧密，为了共同的科研目标或科研任务而分工协作、共同承担责任的研究者集合而成[1]。在大科学时代，随着学科交叉融合的日益加深，研究内容的综合化、复杂化程度不断提升，许多科学研究已经无法由个人单独完成，而需要由优势互补、知识结构多样化的成员组成的核心团队来集体完成攻克。因此，核心团队成为推动学术创新研究的中坚力量。挖掘出合作紧密的核心团队有助于推动机构的科研管理，对于人才评价体系建设与高科技人才引进都有重要的意义。

核心团队识别方法大致可分为 4 类：传统的核心团队识别方法、基于关联规则的核心团队识别方法、基于层次聚类的核心团队识别方法和基于社会网络分析的核心团队识别方法[2]。其中，基于社会网络分析（social network analysis，SNA）是应用最广泛的方法，SNA 以数学、图论等理论为基础，结合中心度、密度等指标，凝聚子群分析、核心－边缘结构分析、网络拓扑结构分析等方法，对计量实体间的关联关系进行测度与可视化呈现[3]。该方法通常以作者间的合著关系、文献间的引用关系等为基础，构建科研合作网络，通过合作阈值的调整得到合作更为紧密的团体，再借助网络节点特性指标（如中间中心度、接近中心度等）来辅助判断网络中的重要作者，最后在数据分析的基础上结合人工判读，以增强团队识别结果的可信度。因此，与其他 3 种方法相比，SNA 在方法可用性与易用性，以及科研团体合作关系、关键科研人员的揭示方面存在显著优势，目前已被广泛应用于社会实体关系的研究。邱均平等[4]根据普赖斯定律筛选计量学领域核心作者，采用 n-clique 指数筛选合作团体，结合中心性类指标判断作者的重要程度；李纲等[5]以合著网络、共词网络和作者关键词耦合网络为基础，采用凝聚子群分析和滚雪球的方法来发现核心团队成员、核心团队研究主题和团队相互间的关系，采用合作网络中心性指标

* 本文系中国科学技术信息研究所青年项目“新兴技术识别方法与演化路径研究——以集成电路领域为例”（项目编号：QN2021－08）研究成果之一。
作者简介：高楠（ORCID：0000－0002－9152－9545），馆员，博士研究生；周庆山（ORCID：0000－0001－8290－940X），教授，博士，博士生导师，通讯作者，E-mail：zqs@pku.edu.cn。
收稿日期：2021－07－19 本文起止页码：81－91 本文责任编辑：徐健

来识别核心团队中的学术带头人[6]；于永胜等[7]提出了基于迭代的中间中心度排名的团队领导人识别方法，再结合 2 – clique 方法和滚雪球方法来识别出核心团队其他成员；赵宇翔等[8]采用社会网络分析的方法，在中心度、集群和小世界效应 3 个方面对开放数据竞赛的行动者协作特征进行实证分析；庞弘燊等[9]选取大连理工大学 WISE 实验室作为科研团队的样本，通过构建多值矩阵、合作网络图并结合社会网络分析方法对样本进行科研团队合作紧密度的分析；沈耕宇等[10]通过向量空间模型计算作者间的相似度，来衡量作者之间的合作关系，再通过社会网络分析中的凝聚子群分析方法分析作者合作关系网络；G. Gonzalez-Alcaide 等[11]分析了 Medline 数据库从 1940 – 2009 年收录的有关南美锥虫病文献的作者合著情况，采用社会网络分析方法确定了由 1 750 位作者组成的 148 个研究团队；X. Liu 等[12]通过二值无向网络模型来构建作者合著网络，将 Author Ranks 定义为单个作者在网络中的影响指标，对数字图书馆领域进行了实证分析；T. Krichel 等[13]分别构建了二值合著网络和加权合著网络，来对 RePEc 数据库内的经济学作者进行合著团队识别；X. Wang[14]、J. Kang[15]等分别将 SAO (Subject-Action-Object)、LDA (latent dirichlet allocation) 模型与社会网络分析法相结合，用于潜在合作伙伴的识别。但这类研究着重于对无向网络的分析，忽略了有向网络对于作者合著网络中隐含信息揭示的有效性；过度依靠网络中心性等指标，而忽略了对作者所属机构、作者署名顺序等重要信息的分析。

因此，笔者将采用有向合著网络来替代传统的无向合著网络，此外，为凸显第一作者与通讯作者的重要性，将分别构建一作 – 其他作者有向合著网络、一作 – 通讯作者有向合著网络用于领域核心团队的识别和团队内指导/师承关系的判断。此外，结合二模网络、关键词共现网络，对核心团队的发刊倾向和研究方向进行揭示。

2 数据来源与预处理

2.1 机构规范与作者去重

在 Web of Science 上设置学科为 Library & Information Science (LIS)，时间跨度为 2016 – 2020 年，索引数据库为 SCI-EXPANDED、SSCI、CPCI-S、CPCI-SSH，检索日期为 2020 年 11 月 2 日，共检索到 12 098 篇文献。本文的数据处理由数据清洗、核心区域作者确定、指导/师承关系判断、团队研究倾向分析 4 部分构成，整体的流程框架如图 1 所示：

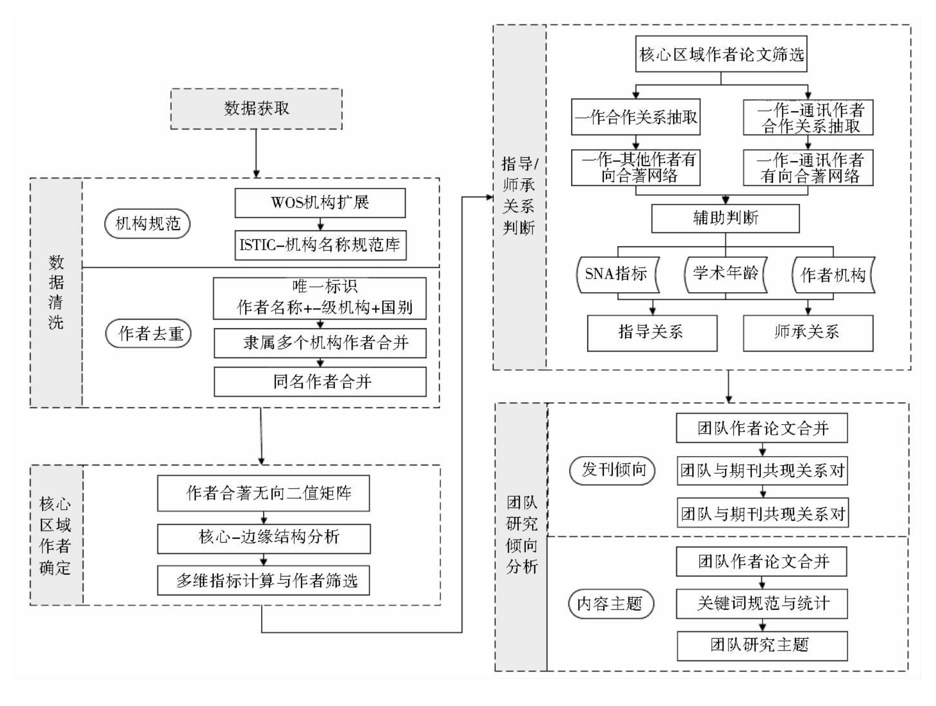


图 1 数据处理流程框架

(1) 机构规范。由于机构名称的表达方式不统一，如全称、简称、俗称等，使得同一个机构会存在多种名称表达方式。而因机构历史沿革导致的名称变化，还会出现同一个机构新旧名称共存的问题。此外，机构间还存在层级关系，因此需要进行机构名称规范。采用 Web of Science 的机构扩展字段进行机构名称的

初步规范,再结合中国科学技术信息研究所自建的“机构名称规范库”进行二次机构名称规范。

（2）作者去重。与机构名称类似,同一作者的名称同样存在多种表达方式,如全称、简称等,采用 Web of Science 的提供的 AF 字段(作者全称),作为作者名称的规范化表达。此外,不同作者间的重名问题,是导致作者成果数量无法准确计算的主要问题之一。在下载的数据集中,发现 Web of Science 提供的 RI(Researcher ID 号)、OI(ORCID 标识符)字段存在大量的数据缺失,且由于 ORCID 可个人免费注册,在注册时提供不同的邮箱,同一作者可同时申请多个 ORCID,准确性有限,因此,笔者采用打标签结合作者合著关系,进行作者同名去重。具体流程包括:①采用“作者名称＋一级机构＋国别”作为唯一标识进行初步统计,得到 31 121 位作者;②将因同一作者所属多个机构,导致的第一机构不一致的作者进行合并。如表 1 所示,对比两篇论文里 M. Chen 挂名的机构名称,判断这两篇论文的作者属同一个人,因此进行作者合并,以“UT 号＋作者次序”为作者赋予唯一标识。

表 1　同一作者判断实例(同一作者挂名多个机构)

论文唯一标识	作者名称	作者次序	机构名称	机构次序	国别	作者唯一标识
WOS:000365550900001	M. Chen	3	State Key Lab Cultivat Base Geog Environm Evolut	1	中国	WOS:0003655509000013
	M. Chen	3	Jiangsu Ctr Collaborat Innovat Geog Informat Reso	2	中国	
	M. Chen	3	Nanjing Normal Univ	3	中国	
WOS:000541354500001	M. Chen	5	Nanjing Normal Univ	1	中国	
	M. Chen	5	State Key Lab Cultivat Base Geog Environm Evolut	2	中国	
	M. Chen	5	Jiangsu Ctr Collaborat Innovat Geog Informat Res	3	中国	

根据周萍、L. Leydesdorff 等的研究[16],在①与②的基础上,将姓名相同且最短合作距离为 2 的作者进行合并。如图 2 所示,在 K. Bystrom 发表的论文中,有两篇均与同一个人 E. E. Isah 合作,因此判断这两篇论文里的 K. Bystrom 属于同一个人,进行同名作者合并后,得到图 2 的 K. Bystrom 合著网络。

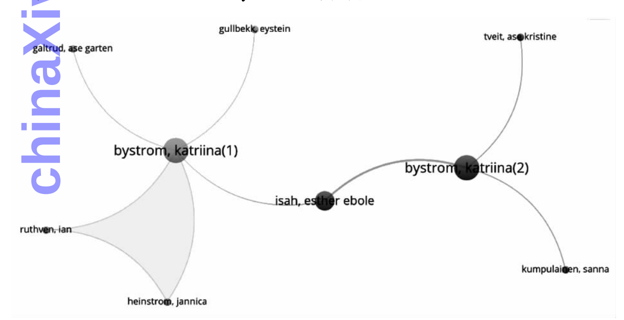


图 2　重名处理前作者合著网络

经过作者重名处理后,共得到 27 221 位作者,发现并处理了 3 900 位同名作者,效果显著。

2.2　相关指标计算

（1）网络节点特性指标:①点中心度反映了一个点与其他点发展交往关系的能力;②接近中心度数值越高表明该节点越接近网络的核心位置,反映结点对于网络中资源的控制能力[17];③中间中心度表示一个点在多大程度上位于图中其他点的中间,是其他点的“中介”;④特征向量中心度的基本思想是,与高得分的节点相连所产生的影响比与低得分的节点相连产生的影响要高[18]。

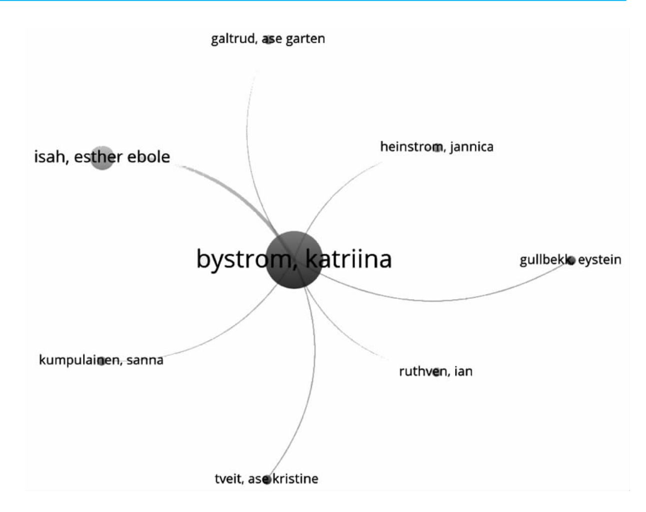


图 3　重名处理后作者合著网络

（2）普莱斯核心作者计算公式如下:

$$P \geq K(p) = 0.749\sqrt{Max(p)} \qquad \text{公式(1)}$$

在公式(1)中,P 代表发文量,K(p)表示入选核心作者候选人的发文量最低值,即某一作者的发文量不小于最低值 K(p)时,该作者才能够入选为核心作者候选人[19]。

（3）H_m 指数计算公式如下:

$$H_m = h + \frac{h}{N_c} \qquad \text{公式(2)}$$

在公式(2)中,N_c 表示作者的总被引次数,由于 $1 \leq 1 + \frac{1}{n} \leq 2$,故 $h \leq h_m \leq 2h$。当某位学者的 $N_c \to +\infty$,

则 h_m→h,因此改学者的学术影响力越大,即总被引频次越高 h_m 与 h 的差值越小[20]。

(4)学术年龄。指作者第一次发文的出版年至今的时长,即该作者开始进行科学研究的时长。在一个合作团队中,学术年龄年长的学者对于学术年龄较轻的学者,或多或少都会进行一些学术上的指导[21]。

3 指导/师承关系判断

3.1 核心区域作者确定

由于本研究的目的是对领域核心团队的挖掘,而同一团队内的作者之间必然存在合著关系,因此,单作者发文的情况不在本研究范围内,需要首先剔除。基于 27 221 位作者间的合作关系进行矩阵构建与可视化,数据量过大且效果不好,因此,剔除作者之间仅合作一次的弱合作关系,基于剩下的合著关系进行网络构建。具体步骤如下:

(1)基于作者间的合著关系构建作者合著无向二值矩阵,仅保留合作次数在 2 次及以上的作者合著关系对(共 3 971 对),结合作者合作次数的分布情况,将合作次数为 2 次的记作"0",超过 2 次的记作"1"。

(2)输入作者合著无向二值矩阵,采用核心 - 边缘结构分析,得到核心区域有 398 个作者,边缘区域有 2 911 个作者。

(3)结合普赖斯核心作者公式、H 指数、H_m 指数、发文量、被引量对核心作者群进行适当调整,具体处理过程如下:①核心区域作者剔除。根据不同指标的作者数量分布,将近 5 年总发文量为 2 篇的 79 位作者,总被引量在 5 次及以下的 19 位作者,H 指数为 1 的 24 位作者,H_m 与 H 指数差值在 0.2 以上的 26 位作者分别筛选出来,合并去重后得到的 99 位作者进行剔除;②边缘区域作者增加。根据普赖斯核心作者公式,核心作者发文量需在 7 篇及以上。因此,将总发文量在 7 篇及以上且总被引量在 25 次及以上的 53 位作者,H 指数 7 次及以上的 18 位作者,H_m 与 H 指数差值在 0.01及以下的 29 位作者分别筛选出来,合并去重后得到 59 位作者增加到核心区域。经过调整,最终确定核心区域作者为 359 人。

3.2 一作 - 其他作者有向合著网络

在一个团队中,第一作者发文量较高的作者通常代表了更高的科研活跃度和科研能力。此外,第一作者及通讯作者相比其他作者而言,对论文的贡献度更高。因此,为凸显第一作者(以下简称"一作")的贡献度,建立一作与其他作者之间有向合著网络。

首先,在数据集中筛选出由 359 位核心作者参与,且非单一作者发文的论文,共计 1 199 篇,涉及 741 位一作。其次,以作者间的合著关系为基础,构建这 741 位作者间的合著关系对。仅计算一作与其他作者之间的合著频次,对非第一作者间的作者合著关系不予计算,使得合著网络中一作的位置更突显,作者间的关系得到了简化。第三,分别建立节点与边的文件。其中,节点文件由作者唯一标识、作者名称、一作发文量构成。边类型设置为有向边,由源节点(其他作者)指向目标节点(第一作者),边权重由作者间的合著论文数量来揭示。整体网络共包含 2 408 个节点,3 162 条边。其中,规模较大的联通网络有 4 个,涉及节点数量分别 795、578、376 和 176 个。

为了对领域核心团队进行更清晰的呈现,提高作者合著阈值到 3,即仅保留一作与其他作者之间合著次数在 3 次及以上的强合著关系,且作者数量在 3 人及以上的团队,删除网络中的孤立点得到图 4。该网络包含 3 个作者及以上的领域核心团队有 25 个,其中,星型的有个 21 个(如 Rousseau Ronald 团队),总线型的有 2 个(如 Glanzel Wolfgang 团队),K 核型的有 2 个(如 Ding Ying 团队)。

3.3 一作 - 通讯作者有向合著网络

在一篇论文的所有作者中,第一作者与通讯作者的贡献度最大。在国际期刊发表学术论文时,多采用"U"型属名方式,即作者署名的首和尾的位置是最重要的,分别是第一作者和通讯作者,其余作者按照贡献度依次进行署名[22]。一般情况下,通讯作者多为项目负责人或一作的导师,为强调其在科研工作中的权利和义务,将通讯作者放在末位,并增加标识。进一步看,如果通讯作者的学术年龄要比一作大,那么通常通讯作者对一作存在学术上的指导行为。所以,为了进一步凸显第一作者与通讯作者在论文发表与网络结构中的重要地位,构建一作与通讯作者间的有向合著网络。

构建过程与一作 - 其他作者有向合著网络类似,但仅计算一作与通讯作者间的合著频次,对于其他作者间的合著关系不予计算,使网络中仅保留了一作与通讯合著的强关系,进一步将合著网络进行简化,有助于探寻作者间的指导、甚至师承关系。分别建立节点与边的文件。其中,节点文件由作者唯一标识、作者名称、一作发文量构成。边类型设置为有向边,由通讯作者(源节点)指向第一作者(目标节点),边权重由作者间的合著论文数量来揭示。

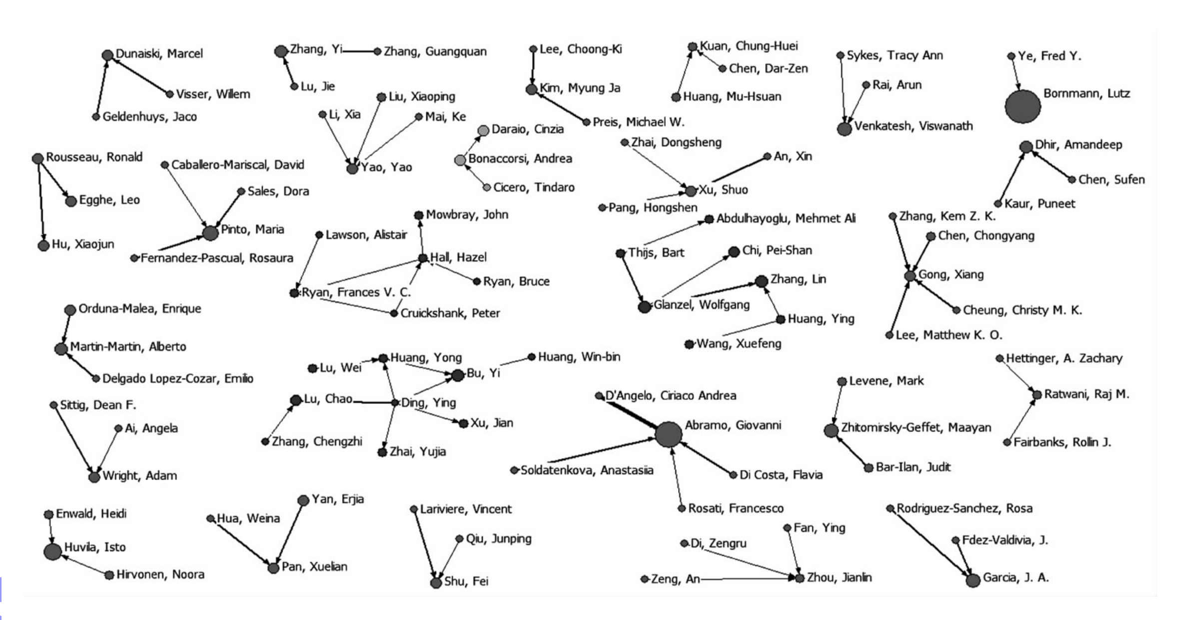


图 4　一作－其他作者有向合著网络

一作与通讯作者间的有向合著网络包含 614 个节点、473 条边，其中，规模最大的连通网络涉及节点数量为 33 个；10 个节点以上的连通网络共有 11 个。网络整体呈现弱连接特点，作者间合作仅 1 次的边数量为 423 条，占总边数的 89%。

为构建领域核心团队的指导/师承关系的判断路径，以团队规模较大且网络拓扑结构具有代表性的 Glanzel Wolfgang 团队（总线型）和 Ding Ying 团队（K 核型）为例。从图 5 可见，Ding Ying 团队的作者合著无向二值网络是典型的 K 核型结构，作者间的联系很紧密，但正因为网络中边数量过多，无法基于此图判断网络中作者间的指导或者师承关系。

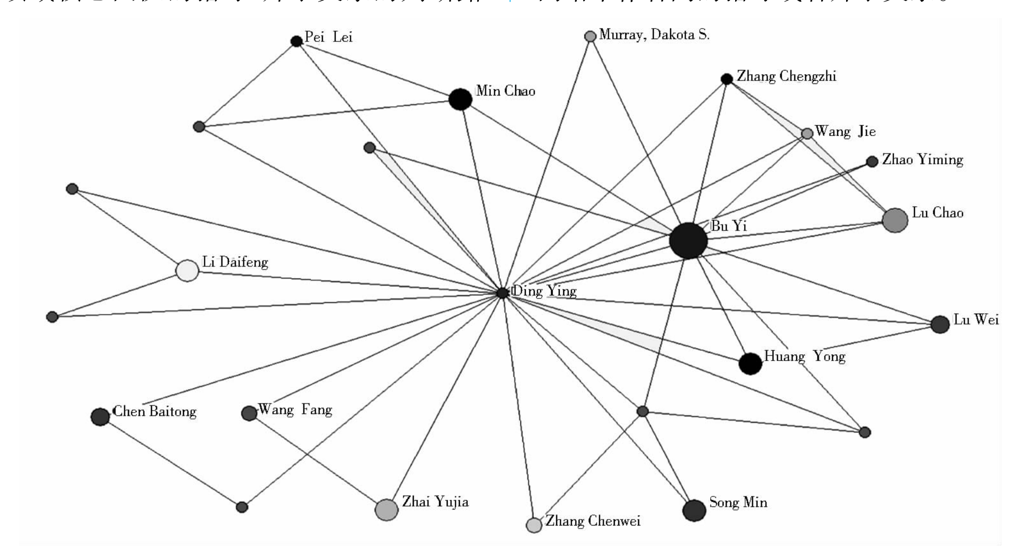


图 5　Ding Ying 团队作者合著无向二值网络

根据上述方法流程，分别构建 Ding Ying 团队的一作－其他作者有向合著网络和一作－通讯作者有向合著网络，由于与 Ding Ying 存在一作与通讯作者关系的仅 Bu Yi 一人，即 Ding Ying 的"一作－通讯作者有向合著网络"仅由一条边构成，因此将其合并到 Ding Ying 的"一作－其他作者有向合著网络"中，用虚线有向边标注。如图 6 所示，图中节点的大小代表作者一作发文量，直线箭头由其他作者指向一作，虚线箭头由通讯作者指向第一作者。相比作者合著无向网络，有向合著网络将 Ding Ying 团队中作者间的关系进行大幅简化，且凸显了一作在网络中的位置，为指导/师承关系的判断提供便利。

从表 2 可见，在 Ding Ying 团队中，度中心度、中间中心度、接近中心度、特征向量中心度最高的 2 位作者

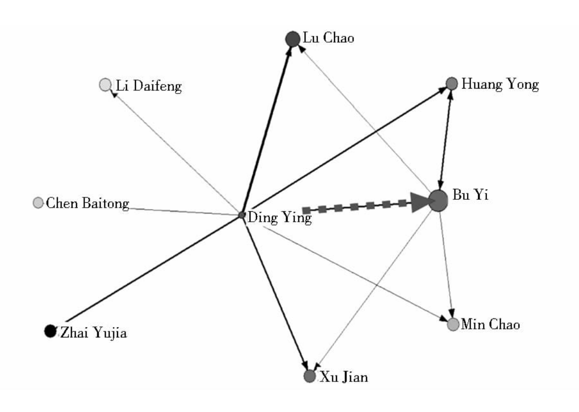


图 6 Ding Ying 团队作者有向合著网络

均是 Ding Ying、Bu Yi，其次是指标比较接近的 Lu Chao、Xu Jian、Min Chao、Huang Yong。这 5 位作者是 Ding Ying 团队作者的有向合著网络中比较重要的节点，与其他人的合作比较频繁。

表 2 Ding Ying 团队网络节点标准化中心性测量

作者	度中心度	接近中心度	中间中心度	特征向量中心度
Ding Ying	100	100	75	83.005
Lu Chao	25	57.143	0	41.502
Zhai Yujia	12.5	53.333	0	22.768
Huang Yong	25	57.143	0	41.502
Xu Jian	25	57.143	0	41.502
Bu Yi	62.5	72.727	10.714	68.303
Min Chao	25	57.143	0	41.502
Chen Baitong	12.5	53.333	0	22.768
Li Daifeng	12.5	53.333	0	22.768

很明显，若作者存在师承关系，在一定时期内二者均属于同一所机构。从表 3 可见，Ding Ying 团队中的作者所属机构中均包含美国印第安纳大学（Indiana Univ）的作者有 Ding Ying、Lu Chao、Huang Yong、Bu Yi、Min Chao。

综合网络节点标准化中心性、作者所属机构，Ding Ying、Lu Chao、Huang Yong、Bu Yi、Min Chao 是师承关系判断的重要作者。从学术年龄看，Ding Ying 的学术年龄是 22 岁，Lu Chao 的学术年龄是 5 岁，Huang Yong 的学术年龄是 3 岁，Bu Yi 的学术年龄是 4 岁，Min Chao 的学术年龄是 5 岁。在一个合著关系中，学术年龄高的学者对学术年龄低的学者或多或少均存在学术指导行为，因此，结合网络节点标准化中心性、作者所属机构、学术年龄，Ding Ying 对 Lu Chao、Huang Yong、Bu Yi、Min Chao 存在指导关系。

综合一作与通讯关系、网络节点标准化中心性、作者所属机构、学术年龄，Ding Ying 与 Bu Yi 在指导关系

表 3 Ding Ying 团队作者所属机构与一作发文量

作者名称	一作发文量	所属机构 1	所属机构 2	所属机构 3
Ding Ying	0	Indiana Univ (USA)	Jilin Univ (China)	Sch Lib & Informat Sci (USA)
Lu Chao	4	Hohai Univ (China)	Indiana Univ (USA)	Nanjing Univ Sci & Technol (China)
Zhai Yujia	3	Nankai Univ (China)	Tianjin Normal Univ (China)	Wuhan Univ (China)
Huang Yong	3	Indiana Univ (USA)	Wuhan Univ (China)	
Xu Jian	3	Sun Yat Sen Univ (China)		
Bu Yi	7	Indiana Univ Bloomington (USA)	Indiana Univ (USA)	Peking Univ (China)
Min Chao	3	Indiana Univ (USA)	Nanjing Univ (China)	
Chen Baitong	2	Shanghai Univ (China)	Wuhan Univ (China)	
Li Daifeng	3	Sun Yat Sen Univ (China)	Tsinghua Univ (China)	

的基础上，有较高概率存在师承关系，即 Ding Ying 是 Bu Yi 的导师（见图 7）。结合 Ding Ying（https://info.sice.indiana.edu/~dingying/me.html）、Bu Yi（https://buyi08.wixsite.com/yi-bu）的个人学术主页，可印证本研究的判断。

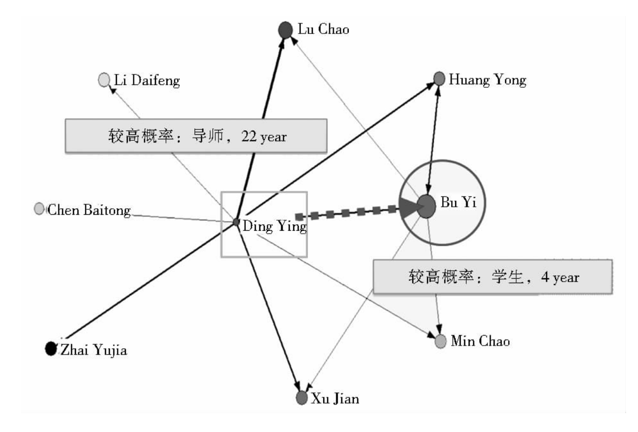


图 7 Ding Ying 团队师承关系判断

与上述流程类似，判断 Glanzel Wolfgang 团队的指导/师承关系。如图 8 所示，判断 Glanzel Wolfgang 有较高的概率是 Chi Peishan、Zhang Lin 的导师，结合 3 位学者的个人学术主页可印证本研究的判断。

从 Ding Ying 和 Glanzel Wolfgang 团队指导/师承关系的判断过程可提炼出通用性规律。首先，在网络拓扑结构为星型的合著团队中，处于中心位置、学术年龄较大且箭头方向为“放射状”的节点有较高概率为导师。因为一方面，这类节点拥有丰富的学术人脉资源，另一方面在论文写作中多处于协助一作研究的角色。如图 9 所示，在“一作－通讯作者有向合著网络”

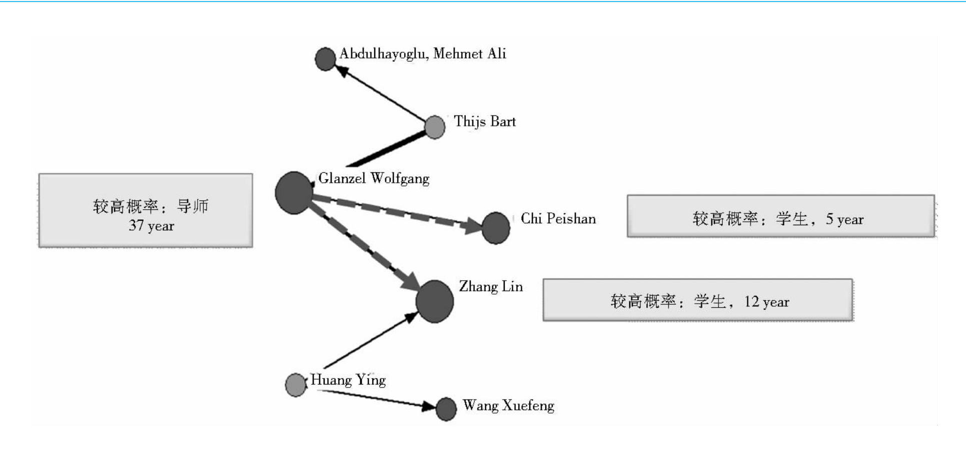


图 8 Glanzel Wolfgang 团队师承关系判断

中，筛选出 6 个网络拓扑结构为星型且规模较大的联通网络。用三角形圈注的节点是有较大概率成为导师/指导者节点，这些节点的有向边均呈现为向外放射状。左下角的连通图虽也呈星型，但中心节点的有向边是向内的聚集状，这类节点在团队中属于高一作发文、科研能力较强的成员，但不能判断出这类节点是否为导师/指导者节点。

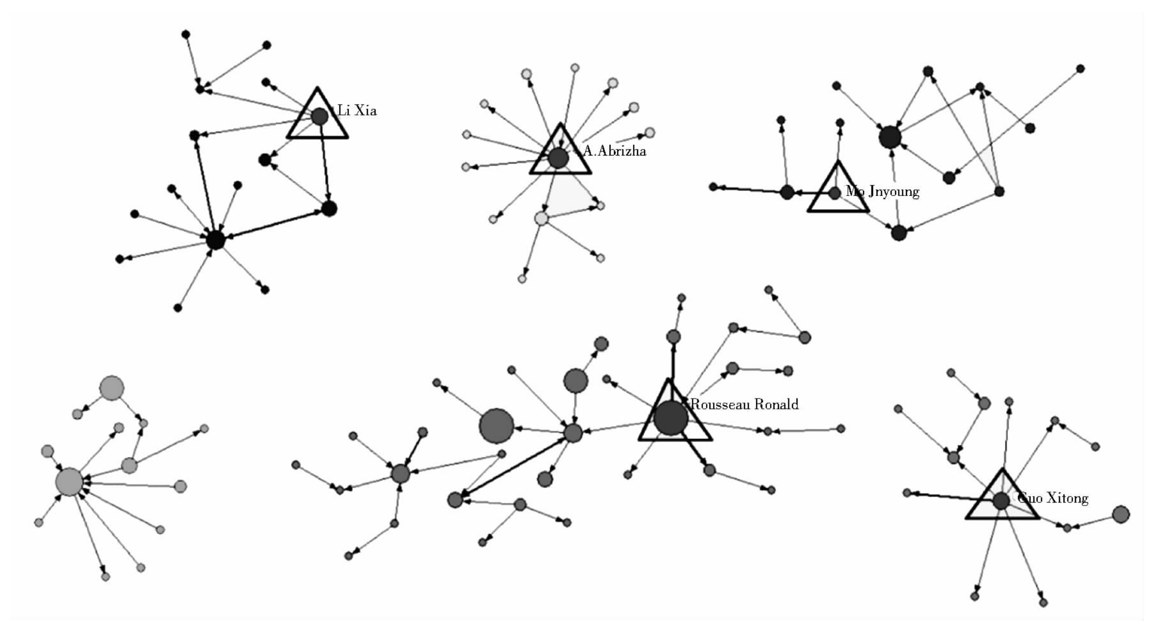


图 9 星型拓扑结构辅助导师节点判断

其次，在“一作 – 其他作者的有向合著网络”中，两个节点若同时具备一作与通讯关系、有相同的所属机构、学术年龄有一定差距，则这两个节点间一定存在指导关系，有较高的概率存在师承关系。若同一合著网络中的节点具备标准化中心性较高、有相同的所属机构、学术年龄有一定差距，可判断出网络节点间有较高概率存在指导关系，是由高学术年龄的学者对低学术年龄的学者进行指导。

4 团队研究倾向分析

4.1 发刊倾向分析

由 3.1 节得到国际情报学领域 3 个作者及以上的领域核心团队有 25 个，为了对团队的发刊倾向进行揭示，需构建团队 – 期刊 2 模网络。首先，合并一个团队所有作者的发文，作为团队的发文量。其次，建立团队与期刊的共现关系对，共计 236 对。第三，分别建立节点与边的文件。其中，节点文件包括两种类型节点，节点类型一由团队唯一标识、团队名称、团队发文量组成，节点类型二由期刊全称、期刊简称、期刊载文量组成（期刊载文量指 25 个核心团队在某刊上近 5 年的总发文量）。边类型设置为有向边，由团队（源节点）指向期刊（目标节点），边权重由团队在期刊的发文量来揭示。

为对结果进行更清晰的呈现，采用成分分析法对

网络进行分割，调节边阈值为 3，其中，最大的联通网络如图 10 所示。该子网络由 20 个节点组成，包括 5 个期刊节点和 15 个团队节点。边粗细代表团队在期刊上的发文量，至少发文 3 篇。5 种期刊聚集了情报学领域 15 个核心团队，这些团队比较偏向于计量类的研究，期刊载文量最多 3 种期刊分别是 *SCIENTOMETRICS*、*J INFORMETR* 和 *J ASSOC INF SCI TECH*。

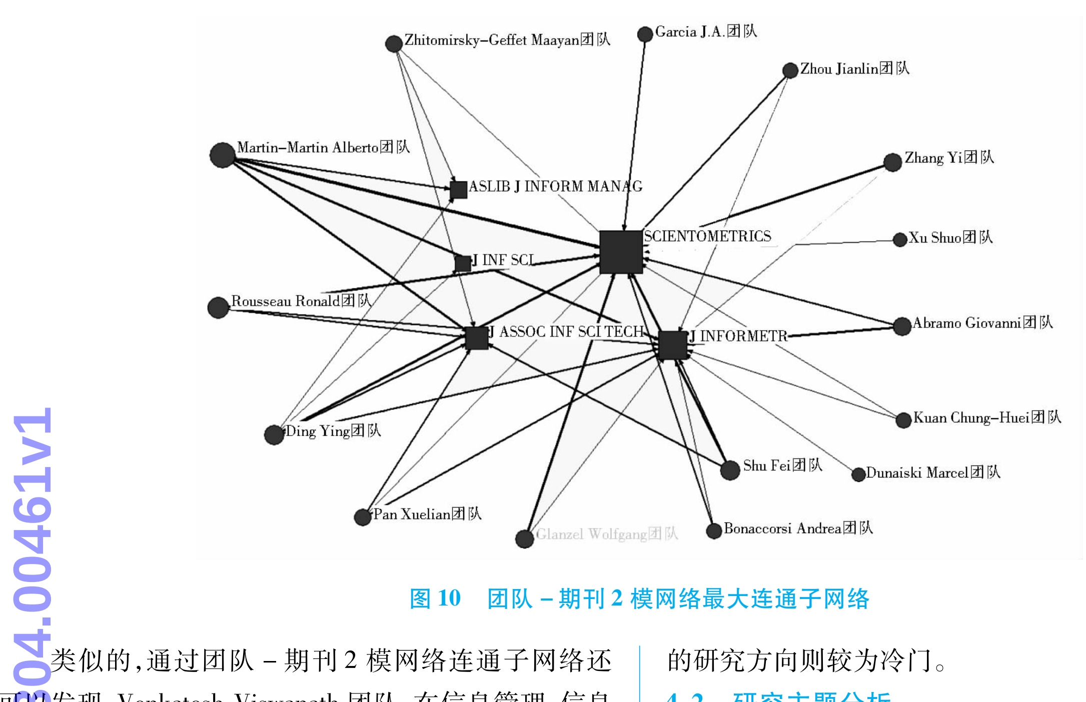


图 10　团队 - 期刊 2 模网络最大连通子网络

类似的，通过团队 - 期刊 2 模网络连通子网络还可以发现，Venkatesh Viswanath 团队，在信息管理、信息系统类期刊发文较多，这些期刊包括 *MIS QUART*、*J ASSOC INF SYST*、*INFORM SYST RES*、*J MANAGE INFORM SYST*；Pinto Maria 团队的发文集中于 *J LIBR INF SCI* 等期刊；Yao Yao 团队集中于地理信息系统领域的期刊 *INT J GEOGR INF SCI*，该团队的研究领域与其他研究相比，具有特殊性，通过这一分析也可以看出，大部分团队的研究主题集中于领域热门方向，有些团队的研究方向则较为冷门。

4.2　研究主题分析

为分析国际情报学领域核心研究团队的研究方向，需对每个团队的发文主题进行揭示。首先，合并一个团队所有作者的发文，作为团队的发文集。其次，从团队发文集中抽取作者关键词，在进行关键词规范后，统计关键词词频，以高频关键词组合反映团队的研究主题，如表 4 所示：

表 4　国际情报学领域核心团队研究主题

序号	团队	高频关键词组合	团队发文量/篇
1	Bonaccorsi Andrea 团队	文献计量学 | Altmetrics | 引文分析 | 评价	20
2	Hall Hazel 团队	数字素养 | 知识管理 | 求职就业 | 社会资本	18
3	Ding Ying 团队	引文分析 | 文献计量学 | 科学计量学 | 主题评价 | 社会网络分析	61
4	Glanzel Wolfgang 团队	引文分析 | 特征值与量表 | 文献计量学 | 混合聚类	42
5	Wright Adam 团队	电子病历 | 临床决策支持 | 电子处方 | 病人安全 | 健康信息技术	57
6	Venkatesh Viswanath 团队	企业系统 | 技术采用 | 工作满意度 | 项目管理	51
7	Zhang Yi 团队	文献计量学 | 文本挖掘 | 跨学科 | 引文分析 | 专利分析	38
8	Rousseau Ronald 团队	文献计量学 | 人文学科 | 评价 | 跨学科 | 专著	67
9	Abramo Giovanni 团队	文献计量学 | 评价 | 科学计量学 | 作者合著 | 生产率	45
10	Huvila Isto 团队	考古学 | 信息工作 | 信息行为 | 老年人 | 信息服务	40
11	Ratwani Raj M. 团队	电子病历 | 技术易用性 | 健康信息技术 | 电子病历 | 病人安全	9
12	Xu Shuo 团队	集群标签 | 协同影响 | 主题建模 | 实体识别	9
13	Shu Fei 团队	文献计量学 | 引文分析 | 国际合作 | 评价 | 科学政策	51
14	Yao Yao 团队	地理空间大数据 | 城市土地利用 | 细胞自动机 | 深度学习	20
15	Pinto Maria 团队	信息素养 | 认知 | 态度 | 移动技术 | 问卷调查	15

（续表4）

序号	团队	高频关键词组合	团队发文量/篇
16	Dhir Amandeep 团队	青少年丨社交媒体丨用户意图与满意度丨数字照片共享	12
17	Gong Xiang 团队	社交媒体丨自我表露丨冲动购物丨游戏成瘾丨情感分析	23
18	**Martin-Martin Alberto 团队**	**引文分析丨Altmetrics丨文献计量学丨评价丨科学计量学丨网络计量学**	**116**
19	Zhou Jianlin 团队	引文网络丨预测丨PageRank丨科学合作	13
20	Kim Myung Ja 团队	智慧旅游丨行为意图丨动机理论丨主观幸福感	11
21	Pan Xuelian 团队	软件引用丨学术交流丨文献计量丨信息计量学丨科学计量学丨自然语言处理	26
22	Garcia J. A. 团队	同行评议丨审稿人丨作者丨编辑丨偏见	11
23	Zhitomirsky-Geffet Maayan 团队	文献索引丨引文分析丨h 指数丨本体论丨数字人文丨知识组织	27
24	Kuan Chung-Huei 团队	引文分析丨耦合分析丨国际合作丨评价丨专利引用	14
25	Dunaiski Marcel 团队	评价丨排名评估丨作者影响指标丨引文分析	4

在统计团队发文量时，发现一个异常值。Martin-Martin Alberto 团队 5 年发文量高达 116 篇，远高于其他团队。在该团队中，M. Thelwall 在 2016－2020 年的一作发文量高达 62 篇，但大部分都是单独作者发文，与其他人的合作强度不高。这类作者属于个人科研能力很强，但与其他作者合作发文不多。类似的作者还有 L. Leydesdorff（一作发文 20 篇）、L. Bornmann（一作发文 41 篇）。

为了按照研究主题对国际情报学领域核心团队进行分类，需要对全领域近 5 年的 12 098 篇论文进行主题聚类分析。首先，抽取领域文献集中的作者关键词，对关键词进行规范。其次，以文献间的关键词共现关系为基础，构建关键词共现矩阵。再者，调节共现强度阈值为 44，共有 80 个关键词参与聚类，采用 modularity 算法进行主题聚类，如图 11 所示：

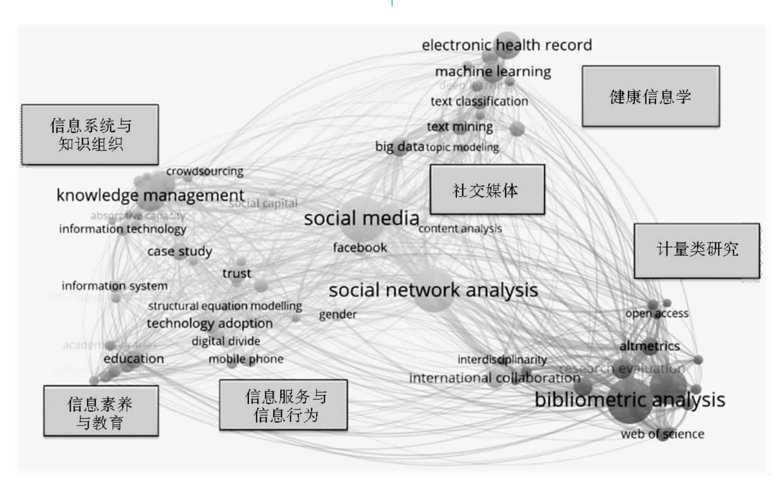


图 11　2016－2020 年国际情报学领域核心研究团队主要研究方向

在图 11 的基础上，将国际情报学领域核心团队进行分类，得到表 5。除计量类研究方向外，其余 5 个研究方向的领域核心团队数量均在 2 到 3 个，竞争程度相对要低一些。

5　结论

笔者主要从核心团队遴选、团队师承关系判别和团队研究方向 3 个方面对领域核心团队进行分析挖掘研究。

笔者首先捋清了领域核心团队遴选流程，包括：①数据前期规范，包括机构规范与作者去重；②构建作者合著无向二值矩阵，采用核心－边缘结构分析法划分核心区域作者，结合多维指标对核心区域作者进行适当调整；③基于核心区域作者的所有发文，构建一作－其他作者有向合著网络，调节作者合著阈值直至团簇结构最清晰，得到领域核心研究团队。在此过程中，本文的创新之处包括：①采用打标签结合作者合著关系，进行作者同名去重；②舍弃传统作者合著无向网络，为凸显一作贡献度，简化网络结构，采用一作－其他作者有向合著网络进行核心团队的识别。

其次，在进行团队师承关系判别分析时，本文的创新之处在于将作者合著的关系强度进行区分，进一步

表 5 国际情报学领域核心团队研究方向分类

研究方向	核心团队		
信息系统与知识组织	Venkatesh Viswanath 团队	Xu Shuo 团队	Yao Yao 团队
信息素养与教育	Pinto Maria 团队	Hall Hazel 团队	
信息服务与信息行为	Huvila Isto 团队	Kim Myung Ja 团队	
社交媒体	Dhir Amandeep 团队	Gong Xiang 团队	
健康信息学	Wright Adam 团队	Ratwani Raj M. 团队	
计量研究	Bonaccorsi Andrea 团队	Ding Ying 团队	
	Glanzel Wolfgang 团队	Pan Xuelian 团队	
	Rousseau Ronald 团队	Abramo Giovanni 团队	
	Garcia J. A. 团队	Martin-Martin Alberto 团队	
	Zhou Jianlin 团队	Kuan Chung-Huei 团队	
	Dunaiski Marcel 团队	Shu Fei 团队	Zhang Yi 团队

凸显了一作与通信作者间的强关系。同时,引入“学术年龄”的概念,为作者之间指导/师承关系的判断提供重要依据。此外,通过典型案例分析提炼了通用性规律,包括:①在“外射星型”拓扑结构的合著团队中,处于中心位置节点有较高概率为导师/指导者;②在“内聚星型”拓扑结构的合著团队中,无法判断中心位置节点是否为导师/指导者节点,但可以判断位于网络中心的作者科研能力较强;③在“一作－其他作者的有向合著网络”中,两个节点若同时具备一作与通讯关系、有相同的所属机构、学术年龄有一定差距,则这两个节点间一定存在指导关系,有较高的概率存在师承关系。

再者,笔者分析了核心团队的期刊发文及研究领域。从团队－期刊 2 模网络,可以清晰地揭示各核心团队的发刊倾向,甚至是研究方向。*SCIENTOMETRICS*、*J INFORMETR*、*J ASSOC INF SCI TECH*、*ASLIB J INFORM MANAG*、*J INF SCI*,这 5 种期刊聚集了情报学领域最多数量的核心研究团队。此外,对于团队的研究方向,笔者分别从粗、细两个维度进行揭示。在细维度层面,对国际情报学领域 25 个核心研究团队的具体研究方向进行列表呈现;在粗维度层面,基于领域近 5 年的 6 个主要研究方向包括:信息系统与知识组织、信息素养与教育、信息服务与信息行为、社交媒体、健康信息学、计量类研究,对 25 个核心团队的研究方向进行归类划分。

本研究以可视化的形式展现了作者间的有向关联关系,凸显了团队中的重要作者,一方面可为基于合著的科研评价提供辅助判断,另一方面也可为信息系统中的关联作者检索提供便利,还可拓展应用于作者互引分析。此外,本文研究结果还便于其他领域学者或新进入图情领域的学者迅速了解 LIS 领域的主要国际情报研究团队与主要研究方向,为加强本领域的合作交流提供一定的参考。但本研究仍存在不足之处,如对于作者间合作的强度采用绝对数值进行揭示,而事实上,与领域普通学者合作和与领域学术权威合作相同数量论文,论文产生的影响力必然不同。后续,笔者将考虑为作者学术影响力赋予一定的权值,采用作者合著相对数值而非绝对数值,对作者间的合作关系与强度进行揭示。

参考文献:

[1] 陈春花, 杨映珊. 基于团队运作模式的科研管理研究[J]. 科技进步与对策, 2002, 19(4): 79－81.

[2] 郭怡婷,邓胜利. 用户关系网络研究综述[J]. 数字图书馆论坛,2014(8):59－65.

[3] OLIVEIRA M, GAMA J. An overview of social network analysis [J]. Wiley interdisciplinary reviews data mining & knowledge discovery, 2012, 2 (2) : 99－115.

[4] 邱均平,伍超. 基于社会网络分析的国内计量学作者合作关系研究[J]. 图书情报知识,2011(6):12－17.

[5] 李纲,李春雅,李翔. 基于社会网络分析的科研团队发现研究[J]. 图书情报工作,2014,58(7):63－70,82.

[6] 李纲,刘先红. 基于合作网络中心性指标的科研团队学术带头人识别研究[J]. 科技管理研究,2016(8):127－132.

[7] 于永胜,董诚,韩红旗,等. 基于社会网络分析的科研团队识别方法研究——基于迭代的中间中心度排名方法识别科研团队领导人[J]. 情报理论与实践,2018,41(7):105－110.

[8] 赵宇翔,仝冲,张妍. 数字人文开放数据竞赛中行动者协作特征研究[J]. 图书馆论坛,2020,40(11):56－67.

[9] 庞弘燊,方曙,杨波,等. 科研团队合作紧密度的分析研究——以大连理工大学 WISE 实验室为例[J]. 图书情报工作,2011,55(4):28－32,99.

[10] 沈耕宇,黄水清,王东波. 以作者合作共现为源数据的科研团队发掘方法研究[J]. 现代图书情报技术,2013(1):57－62.

[11] GONZALEZ-ALCAIDE G, PARK J, HUAMANI C, et al. Scientific authorships and collaboration network analysis on Chagas disease: papers indexed in PubMed (1940－2009)[J]. Revista do instituto de medicina tropical de so paulo, 2012, 54(4):219－228.

[12] LIU X, BOLLEN J, NELSON M, et al. Co-authorship networks in the digital library research community[J]. Information processing & management, 2005, 41(6):1462－1480.

[13] KRICHEL T, BAKKALBASI N. A Social network analysis of research collaboration in the economics community[J/OL]. Journal of information management & scientometrics to appear, 2006, 3

(1). [2021－07－18]. http://core. ac. uk/download/pdf/11879617. pdf.

[14] WANG X, FU Y, HUANG Y, et al. Identifying R&D partners using SAO analysis: a case study of dye-sensitised solar cells[J]. International journal of technology echnology management, 2019, 81(1/2):70－93.

[15] KANG J, LEE J, JANG D, et al. A methodology of partner selection for sustainable industry-university cooperation based on LDA topic model[J/OL]. Sustainability, 2019, 11(12). DOI: 10. 3390//su11123478.

[16] 周萍, LEYDESDORFF L, 武夷山. 中国科技期刊引文环境的可视化[J]. 中国科技期刊研究,2005,16(6):773－780.

[17] 郝治翰,陈阳,王蒲生. 科研合作网络中心性与学术影响力——以Science(2000－2018)为样本[J]. 图书馆论坛,2020,44(4):79－88.

[18] 刘阳. 基于社会网络分析图书情报领域的作者合著关系研究[J]. 四川图书馆学报, 2019(2): 97－100.

[19] 姚雪,徐川平,李杰,等. 基于普赖斯定律和二八定律及在线投稿系统构建某科技期刊核心作者用户库[J]. 编辑学报,2017,29(1):64－66.

[20] 龚舒野. 基于h指数和hm指数的《情报科学》核心作者分析[J]. 情报科学, 2013,31(1):82－85,95.

[21] BU Y, DING Y, LIANG X, et al. Understanding persistent scientific collaboration[J]. Journal of the Association for Information Science & Technology, 2018, 69(3):438－448.

[22] 雷雪,王立学,曾建勋. 作者合著有向网络构建与分析[J]. 图书情报工作,2015,59(5):94－99.

作者贡献说明:
高楠:研究思路设计、数据处理与图表制作、全文撰写;
周庆山:研究思路讨论与修改、论文修改。

Core Team Mining Research Based on Directed Co-Authorship Network ——Taking the Field of LIS as an Example

Gao Nan[1,2] Zhou Qingshan[2]

[1] Institute of Scientific & Technical Information of China, Beijing 100038

[2] Department of Information Management, Peking University, Beijing 100871

Abstract: [**Purpose/significance**] In order to facilitate the introduction of institutional talents and establish a standardized process for the identifying the domain core team, this study researches the core team in the field from four aspects: author identification in core areas, discrimination of team guidance relationship, team publication tendency and team research direction. [**Method/process**] The selection process of the core team included: firstly, carrying out the preliminary data specification; secondly, the core region authors were divided based on undirected binary matrix and core-edge structure analysis, and they were adjusted appropriately combined with multi-dimensional index; finally, the core research teams were identified based on the directed weighted network between the first author and the other authors. [**Result/conclusion**] On the basis of clarifying the basic process of discriminating the team guidance relationship, the following general rules are refined: ①In a co-authoring team of the "outer star" topology, the node at the center has a higher probability of being a mentor, while authors of the "cohesive star" topology whose co-authoring team is located in the center of the network have strong scientific research capabilities; ②In a directed weighted network between the first author and the other authors, if two nodes have the relationship between the first author and the corresponding author at the same time, the same affiliated institution, and a certain difference in academic age, there must be a guidance relationship, and a high probability of a mentoring relationship between these two nodes.

Keywords: social network analysis directed weighted network co-authoring core-edge structure